

Weekly Report

1. 概述

本周我主要的时间花在构思论文的 idea。我完成了以下工作：

- (1) 提交了一份 idea evaluation 给陈老师
- (2) 根据老师指出的我的 idea evaluation 存在的不足继续完善论文 idea

2. 论文 idea 的完善

目前我论文的想法是用迁移学习来解决我之前提出的地图匹配方法（数据驱动的）泛化性不足的问题。

首先地图匹配和一般的迁移学习问题不同，一般的迁移学习问题能够大致知道目标域（domain）的特点，如用文具（源域）分类数据辅助猫狗（目标域）分类。而地图匹配的域（domain）是某路网和轨迹的笛卡尔积：地图匹配域 = 路网 × 轨迹。

为了完成这样的迁移工作，我的想法是用蒙特卡洛法估计地图匹配的目标域。用聚类降低数据的数量。再把源域（已有的路网-轨迹数据）和目标域向量化，并降维至二维空间。把这个二维空间可视化出来，让人类从视图中判断迁移数据集与目标域的相似程度。从数据集中选择人认为最相似的数据，加入训练集中重新训练网络。图 1 是该 idea 的 pipeline。详细的 idea evaluation 我会在近日内发送给老师。

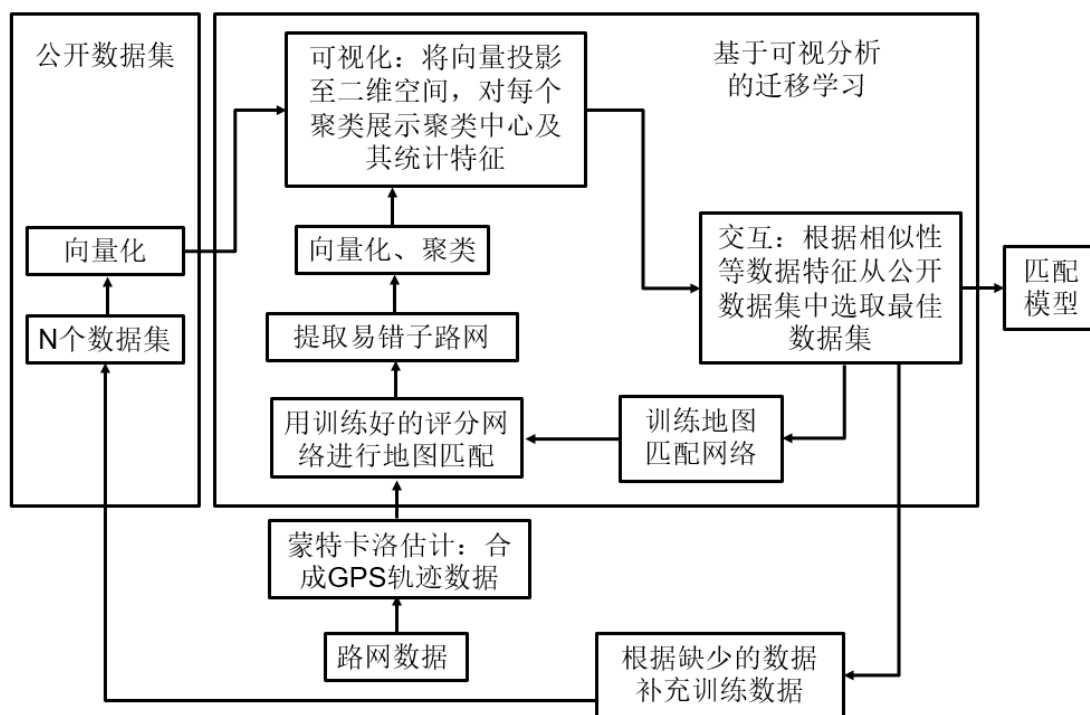


图 1 基于可视机器学习的地图匹配算法

根据老师的反馈，我发现自己的脑中对于迁移学习、蒙特卡洛估计、散点图的特征表达的研究认识得不准确。因此我最近在简单地调研这三个方面的研究。

3. 文献阅读

《Manifold: A Model-Agnostic Framework for Interpretation and Diagnosis of Machine Learning Models》

本文提出了一种通用的机器学习模型的可视化比较方法。

《RuleMatrix: Visualizing and Understanding Classifiers with Rules》

本文提出了一种机器学习模型的规则的理解的交互式可视化视图, 给仅有少量机器学习专业知识的领域专家使用。领域专家可通过使用该交互式可视化视图来判断机器学习模型的好坏。

《SEQ2SEQ-VIS : A Visual Debugging Tool for Sequence-to-Sequence Models》

论文内容如标题所示。

《VIS4ML: An Ontology for Visual Analytics Assisted Machine Learning》

本文提出了一种本体。

《Visual Analytics for Topic Model Optimization based on User-Steerable Speculative Execution》

视觉文本分析、主题模型算法被广泛地作为一个处理过程去有效地分割文档数据集为主题相关的组。本文则提出一种主题模型最优优化可视分析方法来用户可操控地推理和执行。

《Bandit based monte-carlo planning》2000 次引用

本文提出了一种有选择（剪枝）策略的蒙特卡洛模拟方法。

《Efficient selectivity and backup operators in Monte-Carlo tree search》856 次引用

本文首次提出了一个新的框架将树搜索和蒙特卡洛估计结合在一起, 用树搜索方法改进蒙特卡洛估计 GO 游戏的结果。

《A survey of transfer learning》

本文主要调查了适用于 data mining tasks 的 homogeneous transfer learning, heterogeneous transfer learning 和 negative transfer 的内容。

《Learning and Transferring Mid-Level Image Representations using Convolutional Neural Networks》

本文提出了用卷积神经网络学习和迁移中间等级的图像表达。在看文章之前我觉得本文用到的方法就是在 CNN 训练中俗称 fine-tune 的技术。看完本文后, 证实我的观点是对的。本文的核心思想是 CNN 可以当作中间级图像表达的提取器, 只训练最后几层。

《Transfer learning using computational intelligence: A survey》

本文调查了使用计算智能的迁移学习。本文主要提出了四个方面的迁移学习: (1) 基于卷积神经网络的迁移学习, (2) 基于贝叶斯的迁移学习, (3) 基于模糊理论的迁移学习, 此外还提出了基于计算智能的迁移学习的应用。

《Automated Analytical Methods to Support Visual Exploration of High-Dimensional Data》

提出了有分类的散点图质量度量方法。

《Graph-Theoretic Scagnostics》

提出了一种基于图论的散点图质量度量方法。我觉得可以综合这两篇论文中提出的方法得到我想要的数据的 representation。

4. 时间安排

星期	任务	Duration
周一至周五	构思论文 idea	9:00 - 12:00 和 14:00 - 22:30, 共 12 小时

周六	构思论文 idea	12:00 – 22:30
周日	构思论文 idea	8:00-22:30

Work Time: Above 50 hours